

A MÉRÉS-ÉRTÉKELÉS ÚJ TENDENCIÁI: A PAPÍR ÉS SZÁMÍTÓGÉP ALAPÚ TESZTELÉS ÖSSZEHASONLÍTÓ VIZSGÁLATAI ÁLTALÁNOS ISKOLÁS, ILLETVE FŐISKOLÁS DIÁKOK KÖRÉBEN

Csapó Benő, Molnár Gyöngyvér*, Pap-Szigeti Róbert** és R. Tóth Krisztina****

**SZTE Neveléstudományi Intézet, MTA-SZTE Képességkutató Csoport*

***Kecskeméti Főiskola GAMF Kar, SZTE Neveléstudományi Doktori Iskola*

**** SZTE Neveléstudományi Doktori Iskola, Oktatáselméleti Kutatócsoport*

Az oktatási rendszer fejlődése mind több értékelő információt igényel, a tanórai folyamatoktól a közoktatás egészének irányításáig egyre több visszajelzésre van szükség a folyamatok hatékonyabb irányításához. Az utóbbi években megjelent információs-kommunikációs technológiák újabb gyors és költség-hatékony lehetőségeket kínálnak az adatfelvétel, az adatelemzés, és az eredmények közvetítésének megvalósítására. A számítógép alkalmazása nemcsak leegyszerűsíti a tesztelés folyamatát, hanem lehetővé teszi újszerű kérdéstechnika alkalmazását, a multimédia használatát és az adaptív tesztelési megoldásokat is. Figyelembe véve a számítógépes tesztelés kimeríthetetlen lehetőségeit, valószínűsíthető, hogy belátható időn belül ki fogja szorítani a hagyományos, papír-ceruza tesztelést (l. *Kozma, 2009*). A tanulmány keretében áttekintjük a mérés-értékelés legújabb nemzetközi tendenciáit, a formatív, longitudinális és online mérés-értékelés komplex rendszerbe történő integrálásának lehetőségét. Empirikus vizsgálatok segítségével számszerűsítjük a médium közvetítő szerepének hatását általános iskolás, illetve főiskolás diákok körében; továbbá a tanárok és diákok véleményének felhasználásával beszámolunk egy online adatfelvétel tapasztalatról.

A számítógépes tesztelés szerepe a pedagógiai értékelés kiteljesítésében és a tanulás individualizálásában

Az új információs-kommunikációs eszközök akkor segíthetik hatékonyan az oktatás fejlesztését, ha nem a megjelenő technikákhoz keressük a felhasználás lehetőségeit, hanem az oktatás valós problémáinak megoldásában alkalmazzuk azokat. A pedagógiai értékelés, a tesztelés tipikusan olyan terület, amelyen a gyakorlat igénye váltja ki az információs-kommunikációs eszközök alkalmazását. A modern tömegoktatás egyik legnagyobb ellentmondása abból fakad, hogy miközben a tanulók nagyon sokfélék, tanításuk többnyire azonos módon történik. Ennek az ellentmondásnak a feloldását célozza meg a tanulás individualizálása, személyre szólóvá tétele. Az egyéni ütemben való haladáshoz pedig mindenekelőtt a kellően gyakori, részletes és konkrét értékelésre van szükség. A megfelelő keretekbe integrált számítógépes online tesztelés jelentős segítséget nyújthat a tanulás individualizálásához. A lehetőségek kiteljesítéséhez az értékelés három progresszív vonulatának integrálására van szükség: az értékelés formatív, diagnosztikus funkcióinak előtérbe helyezésére, a longitudinális adatfelvételre és a

gyakori, flexibilis adatgyűjtésre, amit csak a technikai eszközök alkalmazásával lehet megoldani.

Az elmúlt évtizedekben a pedagógiai értékelés mindenekelőtt a makro-folyamatokban ért el látványos eredményeket. Rendszeressé váltak a nemzetközi felmérések, a négyévenként lebonyolított TIMSS és a hároméves ciklussal dolgozó PISA lehetővé teszi, hogy a részvevő országok folyamatosan összehasonlítsák saját iskolarendszerük teljesítményét más országokéval, és így felismerjék egyes oktatáspolitikai beavatkozások szükségességét, vagy visszajelzéseket kapjanak azok hatásairól. A nemzeti felmérések sok országban, így nálunk is évenkénti gyakoriságúvá, és néhány évfolyamon teljes körűvé váltak. Az így összegyűjtött információk is elsősorban a rendszer-szintű elemzésekre alkalmasak, szélesebb körű hasznosításához további információkra, viszonyítási adatokra van szükség.

Az elszámoltathatóság elveinek alkalmazása az oktatásban is mind határozottabban jelentkező igény. A tanulók teljesítmény-adataiból úgy lehet az iskolában folyó pedagógiai munka hatásaira következtetni, ha sor kerül az iskolán kívüli, családi, társadalmi hatások figyelembe vételére is: a mért eredményekből valamiképp le kell vonni azt a részt, ami nem az iskolának tulajdonítható. Ennek a legmegbízhatóbb módszerei a tanulók két (vagy több) egymást követő felmérése közötti különbségből indulnak ki. Így szükségessé válik a tanulók egymást követő adatainak összekapcsolása, egyéni követése, azaz a longitudinális adatgyűjtés. Az előző mérés eredményeinek ismerete – praktikusán elektronikus adatbázisban tárolása – felkínálja annak lehetőségét, hogy a tanulók ne uniformizált tesztekkel oldjanak meg, hanem mindenki a tudásszintjének megfelelő, számára optimális kihívást jelentő feladatokkal találkozzon. Egy ilyen differenciált rendszer a gyakorlatban csak számítógéppel valósítható meg.

A rendszerszintű értékelés kiteljesedése jelentős hatást gyakorolt a mérések kognitív-tudományos és pszichometriai hátterének fejlődésére is, melynek eredményei az értékelés újabb területein kerülnek alkalmazásra. Napjainkban az oktatás értékelésével foglalkozó kutatók figyelme mindinkább az osztálytermi folyamatok, és a tanulási környezet mikro-szintje felé fordul, melyet három fő fejlődési tendencia jellemez.

- (1) Az egyéni tanulási folyamatok hatékony támogatásának érdekében a visszacsatolásnak gyakorinak, specifikusnak és pontosnak kell lennie, így a szummatív értékelés mellett egyre nagyobb a formatív és a diagnosztikus értékelés jelentősége. A gyakori mérést nagy eszközigénye miatt a hagyományos tesztelési technikákkal nem lehet megvalósítani.
- (2) Az értékelés viszonyítási alapjául szolgáló referencia pont hagyományosan lehet a népesség átlaga (normatív viszonyítás) vagy valamilyen előre megszabott külső szempont (kritérium-alapú viszonyítás). Az egyéni fejlődés hatékony támogatása szükségessé tesz egy harmadik viszonyítási pont megjelenítését, a tanulók korábbi teljesítményét. Így a diagnosztikus értékelés alkalmazásának logikája is elvezetett a tanulók egymást követő mérési eredményeinek összekapcsolásához, a longitudinális adatgyűjtés szükségességéhez.
- (3) Adatokat gyűjthetjük, tárolhatjuk és feldolgozhatjuk számítógéppel, ezáltal a visszacsatolási folyamatok leegyszerűsödnek, sok elemben automatizálhatóvá, kisebb költséggel megvalósíthatóvá válnak, felgyorsulhatnak.

Az említett három területen, a diagnosztikus értékelés, a longitudinális adatgyűjtés és a számítógépes értékelés terén a fejlődés többé-kevésbé egymástól függetlenül zajlik,

mind gyakoribb azonban az e területek közötti szükségszerű kölcsönhatás. A tanulási folyamatok fejlesztésének gyakorlati igényei miatt pedig szükségessé válik ezek komplex rendszerekbe történő integrálása. Figyelembe kell azonban vennünk, hogy az oktatásban minden jelentős változásnak lehetnek nemkívánatos mellékhatásai is, így a szélesebb körű elterjesztés előtt minden mellékhatást gondos vizsgálat tárgyává kell tenni. A tanulmányban bemutatásra kerülő elemzések is ezt a célt szolgálják.

A számítógépes tesztelés

A fent ismertetett faktorok miatt a számítógépes mérés-értékelés a pedagógiai-pszichológiai értékeléssel foglalkozó kutatások egyik leggyorsabban fejlődő területe, a benne rejlő lehetőségek miatt az értékeléssel foglalkozó fő nemzetközi szervezetek (OECD, CITO) rövid időn belül áttérnek a papír-alapú mérésről a számítógépes tesztelésre (OECD, 2007). Az alkalmazott technológia, item formátum és flexibilitás függvényében a számítógépes tesztelés számos típusa különböztethető meg (*Jurecka és Hartig, 2007*), amelyek a papír-alapú teszteléستől való eltérés mértékének függvényében más-más lehetőséget kínálnak, és más-más feltételeket támasztanak. Az áttérés természetes menete, amikor fokozatosan bővítjük a változtatásra kerülő elemek körét.

A papír-ceruza (paper-and-pencil – PP) tesztekéről a számítógépes (computer-based – CB) tesztekre történő áttérés első lépése, amikor a papír alapú tesztek változtatás nélkül számítógépre visszük (*Csapó, Molnár és R. Tóth, 2008*), azaz a teszt itemeit, itemformáit változtatlanul hagyva vizsgáljuk a közvetítő eszköz szerepét, összehasonlítjuk a két különböző közvetítő médian kitöltött tesztek eredményeit. Ebben az esetben a cél olyan számítógépes tesztek létrehozása, amelyek lehetőség szerint minél több elemében egyezik a papír alapú teszt formátumával, azaz a tesztelés menete továbbra is lineáris marad, a feladatok azonos sorrendben jelennek meg minden egyes tesztelt személy előtt (*Csapó, Molnár és R. Tóth, 2008*). Már ezen a szinten is számos kérdés vetődik fel, például számít-e, hogy pontosan ugyanazon feladatokat lássa a tesztelt személy a képernyőn, amiket papír-alapon egy oldalon lát, vagy elegendő, ha egymás után jelennek meg ugyanazon feladatok, biztosítva a visszalépés lehetőségét. Vajon a teszt tulajdonságainak minimális megváltoztatása hatására még ugyanazt a tudást méri a papír alapú és a számítógép alapú teszt, illetve meddig mérik ugyanazt a tudást (*Molnár, 2008*)? Összehasonlíthatóak-e a különböző médiumon felvett teszteredmények? Vajon milyen betűméret optimális a képernyőn, ami azonos hatást kelt az analóg papír alapú teszttel, mennyire befolyásolja a számítógépes teszt megoldójának eredményét számítógépes jártassága?

A papír-ceruza tesztekéről a számítógépes tesztekre történő áttérés egyik következő lépése lehet, amikor a tesztelés továbbra is lineáris marad, de az itemek formátuma változik. A tesztelésbe bevonunk olyan újabb típusú multimédiás itemeket, amelyek megjelenítésre papír-alapon nem lehetséges. A technológia adta lehetőségek nagyobb mértékű kihasználásával alkalmazhatunk például hang, mozgó kép, animáció, szimuláció, interaktív szimuláció elemekkel gazdagított itemeket is.

A papír-ceruza tesztekéről a számítógépes tesztekre történő áttérés egy másik lehetséges egyszerűbb továbblépési esete, amikor megőrizzük az itemek papír-ceruza formátumát, viszont megváltoztatjuk a teszt linearitását. Erre több lehetséges eszköz is rendelkezésre áll. A feladatok véletlenszerű kiválasztásától, az itemek előzetes

csoportosítása után a létrehozott csoportokból randomizált itemválasztáson keresztül egészen az automatikus itemgenerálásig, amikor az adott típusfeladat, például szöveges feladat, mindig más-más változóértékkel (név, szám) jelenik meg. Ezek a tesztelési eljárások nem veszik figyelembe a diák képességszintjét, más változókat figyelembe véve történik az itemek kiválasztása (Molnár, 2008).

A számítógépes tesztelés legmagasabb szintjén a feladatok kiválasztása a vizsgázó korábbi válaszainak függvényében történik, ezáltal a teljesítmények sokkal finomabb felbontását, mérését teszi lehetővé. A tesztelés ezen a szinten megvalósítja a teljes mértékben személyre szabott tesztelést, ami elhanyagolhatóvá teszi annak valószínűségét, hogy mindenki ugyanazon teszten kapná a tesztelés folyamata során. Ennek, azaz az adaptív tesztelés (Computerized Adaptive Testing – CAT) megvalósításához szükséges egy teljes mértékben parametrizált, indexelt és egy azonos nehézségi, illetve képességskálán leírható feladatbank. Az adaptív tesztelés főbb ismérveiről és menetéről részletesebben lásd *Csapó, Molnár és R. Tóth (2008) tanulmányát.*

Feltehető, hogy minél inkább megfeleltethető egymásnak flexibilitásban, feladattípusok, alkalmazott elemek tekintetében a papíron, illetve számítógép segítségével kitöltött tesztek, annál kisebb a médiahatás. Ezt a feltevést azonban konkrét elemzésekkel kell igazolni, és meg kell határozni, milyen mértékűek az említett hatások. A jelen tanulmányban bemutatásra kerülő kutatások ezt a célt szolgálják. Feltehetőleg minél inkább kihasználjuk a számítógép adta lehetőségeket, a számítógép előtt írt és a hagyományos tesztek különböző feladattípusain elért eredmények annál inkább eltérnek egymástól (Csapó, Molnár és R. Tóth, 2008).

A továbbiakban áttérünk a számítógép- és papír-alapú tesztelés eredményeinek összehasonlító vizsgálatára, amit két, egymástól függetlenül zajló, de sok tényezőben egymáshoz hasonló kutatás fényében teszünk meg.

A számítógép- és papír-alapú tesztelés eredményeinek összehasonlító vizsgálat

A kutatás célja

A papír-ceruza tesztelésről a számítógépes tesztelésre való áttérés első lépéseként a teszt formátumának minél több szempont szerint történő megőrzése mellett a papír-alapú teszt digitalizálását javasoltuk. Ez lehetőséget teremt a teszt médiuma befolyásoló szerepének minél pontosabb vizsgálatára, amire korábban hazai kutatások során nem került sor.

A tanulmányban bemutatásra kerülő kutatások célja

- (1) a mérés-értékelés médiuma befolyásoló hatásának különböző változók mentén (itemtípus, kontextus, életkor) történő diagnosztizálása;
- (2) a tanulói teljesítmények közötti azonosságok, illetve különbségek okának feltárására;
- (3) teljesítmény tekintetében a közvetítő médiára érzékenyebb tanulócsoporthoz leírása;
- (4) az online teszteléssel kapcsolatos attitűdök, problémák feltérképezése.

A bemutatásra kerülő empirikus vizsgálatok filozófiája, kutatási kérdései, alkalmazott mérőeszközei hasonlóak, mintájuk és a kutatás formája, tervezése eltérő. A kutatások bemutatása során nem azok sorrendjét vesszük alapul, hanem a mintát alkotó

diákok életkorát. Ennek megfelelően a később, de fiatalabb diákok körében zajló kutatás mintáját, módszereit és eredményeit mutatjuk be elsőként, amit minden esetben az idősebbek körében, de két évvel korábban történt kutatás ismertetése követ.

Minta

A 2008 májusában történő adatfelvételben 5. évfolyamos diákok vettek részt (n=843) 24 település, 34 iskolájából. A minta 53%-a fiú volt. A minta kiválasztása során nem volt cél a reprezentativitás. Ennek ellenére egy azonos korú reprezentatív és a jelen kutatás mintájának szülők iskolai végzettsége szerinti eloszlásában ($\chi^2=7,13$, $p>0,05$) nincs szignifikáns különbség.

A két évvel korábban, 2006 tavaszán zajlott vizsgálatba 268 főiskolai hallgató vett részt. A minta a szülők iskolai végzettsége szerint megfelel az országos eloszlásnak (anya: $\chi^2=4,93$; $p>0,05$), a nemek azonban az intézményi arányoknak megfelelően képviseltek (a férfiak aránya a mintában 87%). Eredményeinket ezért csak az érintett főiskola hallgatóira általánosíthatjuk. Az egyes évfolyamok hallgatói az intézményi arányoknak megfelelően szerepeltek a vizsgálatban. A hallgatók mindegyike nagy óraszámban tanult informatikát, így a médium ismeretlensége valószínűleg nem okoz torzító hatást.

Mérőeszköz

Mindkét kutatás ugyanarra a papír-alapú mérőeszközre alapoz és digitalizálja azt más-más online környezetben (erről részletesebben l. később). A kutatások tervezése során két ok miatt esett a választás *Csapó Benő* által kidolgozott és korábban már széles körben alkalmazott induktív gondolkodás tesztre (2001). Egyrészt az induktív gondolkodás a megismerés egyik leggyakrabban vizsgált területe (*Csapó*, 2001), a gondolkodási képességek között kiemelkedő fontosságú, mivel a tudás, az új ismeretek megszerzésének egy fontos képessége. Másrészt feltételezhető, hogy egy általános kognitív képesség esetében nem történik szignifikáns fejlődés a két, időben egymáshoz közeli adatfelvétel (papír-alapú és számítógépes) között, ami elkerülhetetlen, ha ugyanazon mintán ugyanazon teszt különböző megjelenési formáinak vizsgálatát tervezzük. (Az eredmények alátámasztották ezt a feltételezést.) Az 58 ítemes teszt három résztesztből áll: számanalógiák, számsorok és szóanalógiák. A számanalógiák és számsorok részteszt feleletalkotó itemeket, míg a szóanalógiák feleletválasztó itemeket tartalmaz.

Az induktív gondolkodás teszten felül a fiatalabb életkori mintán történő kutatás során a diákok mind az online adatfelvétel előtt, mind utána kitöltöttek egy-egy kérdőívet. Az első, számítógép használati szokásaikra, a második, a teszteléssel kapcsolatos attitűdjeikre kérdezett rá. A felügyelő tanárok részére készített harmadik kérdőívvel a teszteléssel kapcsolatos tapasztalatokról a kaptunk információt.

A főiskolás hallgatók körében történt kutatás kiegészítő mérőeszközei a szociális készségek és motívumok, illetve a tanulási motívumok mérésére alkalmas kérdőívek voltak. Előbbit *Zsolnai Anikó* és *Józsa Krisztián* (2002), utóbbit *Józsa Krisztián* és *Székely Györgyi* (2004) állította össze.

A mérés lebonyolítása, adatfelvétel

A 2008-ban végzett adatfelvétel során minden egyes tanuló először papíron, majd néhány hét különbséggel számítógépen is megoldotta az induktív gondolkodás fejlettségét mérő tesztet. A teszt megoldására mindkét formátumban 35 perc állt a tanulók rendelkezésére. Ez a típusú kutatás, amikor minden egyes diák mindkét médián megoldja az alkalmazott tesztet nemzetközi viszonylatban is egyedi. Ebben az esetben diákszinten összehasonlítható az alkalmazott médium, közvetítő eszköz befolyásoló hatása. A háttérkérdőívek felvétele mindenki esetében papír alapon zajlott. Egy másik típusú adatfelvételi módszert alkalmaztunk a két évvel korábbi kutatásban.

A 2006-os kutatás során a minta egy részmintája, az évfolyamok és a nemek arányát jól reprezentáló 73 fős részminta hallgatói valamennyi mérőeszközt, mind az induktív gondolkodás tesztet, mind a különböző készségeket, motívumokat vizsgáló kérdőíveket elektronikusan töltötték ki. A minta másik része a hagyományos, papíron történő mérésben vett részt. Mindegyik mérőeszköz kitöltésére azonos idő állt rendelkezésükre függetlenül a teszt médiumától. Ebben a kutatási formában a háttérváltozók tekintetében hasonló diákok eredményeinek összehasonlító vizsgálatára került sor. A kutatás során nemcsak egy kognitív teszten elért eredményeket, hanem az alkalmazott kérdőívekre adott válaszok alakulását is összehasonlíthattuk, ha azokat papír vagy számítógépes alapon töltötték ki a diákok.

Mindkét kutatás során törekedtünk arra, hogy az induktív gondolkodás teszt számítógépre adaptálása során a teszt minél több tulajdonságát megőrizzük. Ez az alkalmazott webes felület függvényében különbözőképpen valósult meg. Mindkét kutatás során megtartottuk a teszt linearitását, azaz minden egyes diák ugyanabban a sorrendben kapta ugyanazokat a feladatokat. A 2008-as kutatás során az alkalmazott webes felület nem tette lehetővé, hogy az egyszerre a képernyőn látható feladatok mennyisége megegyezzen a papír-alapú tesztfüzet egy oldalán található feladatokéval, a tesztfeladatok közötti navigálást egy előre és visszafelé lépést lehetővé tevő navigációs gombbal valósítottuk meg. Ez a feltétel, azaz a képernyőn egyszerre megjelenő és a papír alapú feladatlapon egy oldalon lévő itemek mennyiségének egyezése a 2006-os kutatásban teljesült. A digitalizálás során hasonló módon történt a feleletválasztó itemek átalakítása. A papíron karikázással megvalósítandó válaszadást a számítógépen rádiógomb használatával oldottuk meg, azaz kattintással kellett kiválasztani a helyesnek ítélt megoldást. A feleletalkotó számanalógiák és számsorok esetén a digitalizált tesztváltozat esetében az adatbeviteli ablakok között egerrel vagy tabulátorral kellett továbblépni és begépelni a helyes választ.

A 2008-as kutatásban a számítógépes adatfelvétel interneten zajlott a TAO (Testing Assisté par Ordinateur – számítógépes tesztelés) platformon keresztül (*Plichart, Jadoul, Vandenabeele és Latour, 2004; Farcot és Latour, 2008; Martin, 2008*). A TAO egy nyílt forráskódú szoftver, melyet a Luxemburgi Egyetem és a Centre de Recherche Public Henri Tudor Intézet tervezett és implementált, az SZTE Oktatáseméleti Kutatócsoportja pedig adaptálta és magyarította.

A 2006-os vizsgálat adatfelvétele saját fejlesztésű webes felületen zajlott, ami szintén nyílt forráskódú eszközöket használt fel.

Eredmények

Általános iskolás diákok papír, illetve számítógépen mutatott teljesítményeinek összehasonlító vizsgálata, teszt-, részteszt és itemszintű elemzések

Az induktív gondolkodás teszt reliabilitásmutatója papír (Cronbach- $\alpha=0,90$), illetve számítógépes (Cronbach- $\alpha=0,91$) formátumban jelentősen nem különbözött egymástól. A diákok különböző közvetítő médian mutatott átlagos teljesítménye között szignifikáns korreláció ($r=0,78$, $p<0,01$) van. Az összefüggések szorossága résztesztenként változik, de minden esetben ($p<0,01$) szignifikáns. A leggyengébb kapcsolat a számsorok esetében figyelhető meg ($r=0,42$), míg a legerősebb a szóanalógiák esetében ($r=0,80$).

A diákok papíron ($x=27,2\%$, $sd=14,9\%$) és számítógépen ($x=26,0\%$, $sd=14,0\%$) elért átlagos teljesítménye szignifikánsan ($t=3,6$ $p<0,001$) különbözik egymástól. A résztesztenkénti elemzés alapján megállapítható, hogy a diákok a feleletalkotó itemeket tartalmazó számanalógiák és a számsor részteszten szignifikánsan jobb eredményt értek el tradicionális formátumban, mint online (1. táblázat). A feleletválasztós itemeket tartalmazó szóanalógiák részteszten fordított kép mutatkozik, az online teszten jobban ($x=42,8\%$, $sd=22,2\%$) teljesítettek a diákok, mint papír alapon ($x=40,4\%$, $sd=21,4\%$; $t=-5,27$, $p<0,001$). A legnagyobb médiahatás a számolás igénylő nyílt végű feladatok esetében figyelhető meg.

1. táblázat. A papír- (PPT) és számítógép (CBT) teszten mutatott teljesítmények teszt és részteszt szintű összehasonlítása

		Átlagok eltérése	sd	t	p
CBT _{teljes}	– PPT _{teljes}	-1,17	9,48	3,57	0,000
CBT _{számanalógia}	– PPT _{számanalógia}	-1,62	17,53	2,68	0,007
CBT _{szóanalógia}	– PPT _{szóanalógia}	2,50	13,78	-5,27	0,000
CBT _{számsorozat}	– PPT _{számsorozat}	-4,38	11,88	10,71	0,000

Az itemek médiafüggő viselkedését támasztja alá a közös nehézségi skálára konvertálás eredménye is, ahol megfigyelhető, hogy a feleletválasztós, számolást nem igénylő feladatokat bátrabban oldják meg a tanulók, azok nehézségi indexe csökkent, míg a számolást igénylő feladatok nehézségi indexe nőtt.

Főiskolai hallgatók papír, illetve számítógépen mutatott teljesítményeinek összehasonlító vizsgálata, teszt-, részteszt és itemszintű elemzések

Az induktív gondolkodás teszt papíron felvett változatának reliabilitásmutatója főiskolás diákok esetében is Cronbach- $\alpha = 0,90$, az elektronikus változaté Cronbach- $\alpha = 0,88$. A mutatók hasonlóan a 2008-as kutatásban tapasztaltakhoz jelentős mértékben nem térnek el egymástól.

A vizsgálatban részt vett hallgatók induktív gondolkodásának átlagos fejlettsége (61,9%) megegyezik az érettségit adó intézmények 11. évfolyamosai átlagával (*Csapó,*

2001), a sikertelen tantárgyteljesítések valószínűleg nem e képesség hiányosságaira vezethetők vissza elsősorban.

A minta eredményeinek eloszlása normáloszlás-jellegű. Az itemek – közös nehézségi skálára konvertálva – a szélsőértékeket is figyelembe véve lefedik a hallgatók képességszintjét, bár az itemek nehézség szerinti eloszlása nem egyenletes. A minta értékeléséhez néhány további nehéz item lenne szükséges, az itemnehézség szempontjából a könnyű itemek egy része a főiskolás minta esetén elhagyható (az országos eredményekkel való összevetés miatt azonban érdemes ezeket az itemeket is megtartani). A nehéz itemek többnyire a számsorozatok megoldását jelentik.

A papíron ($x=61,9\%$, $sd=14,7$) és számítógépen ($x=61,6\%$, $sd=15,9$) válaszoló hallgatók eredménye a 2008-as kutatás eredményével ellentétesen a teljes teszten nem tért el szignifikánsan egymástól (2. táblázat). Az egyetlen szignifikáns különbség a számsorozatok kitöltése során jelentkezett: ezen a részteszten a papíron dolgozó hallgatók eredménye szignifikánsan jobb. Ez alátámasztja a 2008-as kutatás során ezen a részteszten tapasztaltakat. Feltételezhető, hogy a természetes módon rendelkezésre álló papírt jobban használták fel a hallgatók a részeredmények feljegyzésére (annak ellenére, hogy számítógépes tesztet kitöltők számára sem tiltottuk a papír használatát).

2. táblázat. A papír- (PPT) és számítógép- (CBT) teszten mutatott teljesítmények teszt és részteszt szintű összehasonlítása

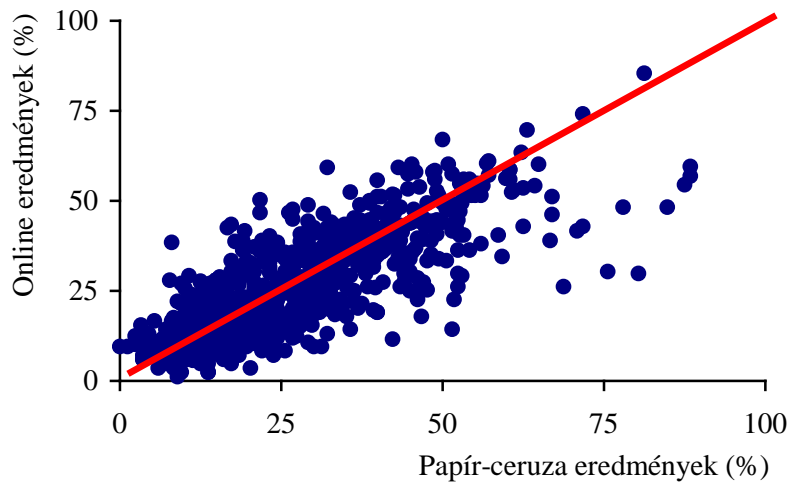
		Átlagok eltérése	Szórás CBT	Szórás PPT	t (d)	p
CBT _{teljes}	– PPT _{teljes}	0,3	14,7	15,9	0,18	0,861
CBT _{számanalógia}	– PPT _{számanalógia}	3,1	27,5	25,1	-0,85	0,395
CBT _{szóanalógia}	– PPT _{szóanalógia}	2,7	12,2	14,1	-1,46	0,146
CBT _{számsorozat}	– PPT _{számsorozat}	-6,9	18,2	22,7	2,321	0,021

A háttérváltozókkal végzett vizsgálatok legfontosabb törekvése, hogy feltárja azokat a faktorokat, amelyek az online tesztelés során pozitívan vagy negatívan érintenek részmintákat. A háttérváltozók közül a továbbiakban, az ötödik osztályos mintát tekintve kiemelten foglalkozunk a nemek, illetve a számítógépes jártasság szerepével. Az ott tapasztaltakkal egybecsengően a főiskolai diákok körében sem mutatható ki szignifikáns teljesítménykülönbség a teljes teszt vonatkozásában a fiúk és lányok teljesítmény között. A 2008-as kutatás eredménye azonban nem igazolja a részteszteken végzett elemzések eredményeit, ahol főiskolai szinten szintén nem mutatható ki szignifikáns különbség az egyes résztesztjelekben.

Általános iskolás diákok online teszteléssel kapcsolatos attitűdjei és eredményei a háttérváltozók tükrében

Az 1. ábra az ötödik évfolyamos tanulók papíron és számítógépen elért eredményeit mutatja. Egy pont egy tanulót reprezentál, a vízszintes tengely a papíron elért, a függőleges tengely az online elért eredményt ábrázolja százalékban kifejezve. Optimális esetben megegyezik a két médiumon elért teljesítmény, ekkor a pont a vastagított vonalon, vagy annak közelében helyezkedik el. Vannak olyan tanulók, akik azonban

távolabb helyezkednek el az elvárt egyenestől; az eltérés okának feltárására a háttérkérdőívben felvett adat alapján keressük a választ.



1. ábra
Papír-ceruza és online tesztek eredményeinek összehasonlítása

A részmintákra irányuló elemzés eredményei szerint a fiúk és lányok átlagos összteljesítménye egyik médium esetében sem különbözött egymástól (3. táblázat). Nemeken belüli bontásban a lányok papíron ($x=27,65\%$, $sd=14,33\%$) szignifikánsan jobban teljesítettek, mint számítógépen ($x=26,04\%$, $sd=13,89\%$; $t=3,5$ $p<0,001$). A fiúk összteljesítménye médiafüggetlen volt.

A résztesztenkénti elemzés alapján megállapíthatjuk, hogy mindkét nem szignifikánsan ($p<0,05$) jobban teljesített online környezetben a 28 feleletválasztós íte mből álló szóanalógiák és rosszabbul a számsorok részteszten. Ez a feladattípusok sajátosságaival magyarázható.

3. táblázat. Nemeken belüli és nemek közötti különbségek a média függvényében

	Papír alapú (PP)	Számítógépes (CB)	Lány	Fiú
	Nemek között		Nemeken belül	
Induktív gondolkodás	n.s.	n.s.	PP>CB; $p<0,05$	n.s.
Számanalógia	n.s.	n.s.	PP>CB; $p<0,05$	n.s.
Szóanalógia	lány>fiú; $p<0,05$		CB>PP; $p<0,05$	CB>PP; $p<0,05$
Számsor	n.s.	n.s.	PP>CB; $p<0,05$	PP>CB; $p<0,05$

Hipotézisünk szerint a diákok médiafüggő viselkedését befolyásolhatja az IKT-val kapcsolatos tapasztalatuk és hozzáállásuk. Az eredmények alapján ezek a mutatók szignifikánsan különböznek egymástól a fiúk és lányok tekintetében. A fiúk általában több időt töltenek el számítógép előtt, több tapasztalatuk van használatában, mint a

lányoknak. De ez az összefüggés nem kizárólagos magyarázó tényezője az eredményeknek, számos faktor befolyásolja a diákok médiafüggő, illetve független viselkedését.

A tanulói aktivitás vizsgálatának eredményei arra engednek következtetni, hogy a hallgatók aktívabbak, több választ adnak online környezetben, mint a hagyományos tesztelés során. Ez magyarázható a két közvetítő médián adott válaszok számával, és a tanulók attitűdjével. Ezek alapján több mint 700 válasszal több született a számítógépes megoldás során, mint a papír-alapú kitöltéskor. Továbbá a diákok 79%-ának jobban tetszett az online változat, mint a papír-ceruza teszt.

Főiskolai hallgatók képességeinek és motívumainak papíralapú és elektronikus mérése

A 2006-ban zajlott kutatás lebonyolításának menete nem tette lehetővé, hogy összehasonlítsuk ugyanazon diákok papír, illetve számítógépes eredményeit, viszont a kutatás keretében más, a diákok szociális készségeit, illetve motívumait mérő kérdőívek is felvételre kerültek. Ebben az esetben is a diákok egy része papíron, másik része számítógépen töltötte ki a kérdőíveket.

A szociális készségeket és motívumokat illetve kognitív motívumokat vizsgáló kérdőív esetén feltételezésünknek megfelelően nem tapasztaltunk különbséget a papíron illetve elektronikusan felvett adatok között. Az oktatás problémáját elsősorban a szociális motívumok, a tanulási énkép alacsony szintje, valamint az elsajátítási motívumoknak a közoktatás végítől valamivel elmaradó fejlettsége jelenti (Józsa, 2007). Az elsajátítási motivációban megjelenő különbségek feltételezésünk szerint jelentős szerepet játszanak a képzésben való megmaradásban, a harmadik évfolyam átlaga mintegy 8 % ponttal magasabb az elsősökénél. A szociális készségek fejlettsége megfelel a közoktatás végi országos helyzetnek.

Vizsgálatunk eredményei egyrészt felhívják a figyelmet arra, hogy a hallgatókkal kapcsolatos sztereotípiák nem minden esetben helytállóak, a hallgatók kognitív fejlettségét jelző induktív gondolkodás átlagos fejlettsége nem marad el számottevően az érettségit szerző tanulókétól. Szükséges lenne ugyanakkor a hallgatói motívumok részletesebb megismerésére, a motívumokat fejlesztő oktatási módszerek kialakítására. Az elektronikus és a papír alapú tesztek eredményének összehasonlítása azt mutatta, hogy a nagyobb mennyiségű részeredmények megjegyzését igénylő feladatokban az elektronikus mérőeszköz (a rövidtávú memória korlátai miatt) alulbecsülheti a képesség fejlettségét. Érdemes lenne eredményeinket a papírhasználatra explicit módon javaslatot tévő, illetve összetettebb feladatokat is tartalmazó vizsgálattal ellenőrizni.

Következtetések

A tanulmányban bemutatott kutatások egyik legfontosabb tapasztalata, hogy a hazai iskolákban az online teszteléshez szükséges alapvető feltételek biztosítottak. Az interneten lebonyolított tesztelés folyamán komolyabb problémák nem merültek fel, ezért ajánlott a számítógépes tesztelés formatív és diagnosztikus mérés-értékelés során történő fokozatos bevezetése. A nagy tételt bíró (high stakes) tesztelés online megvalósításához azonban még további feltételek teljesülése szükséges (Csapó, Molnár és R. Tóth, 2009).

A jelen kutatások eredményei valószínűsítik, hogy az áttérés folyamán még jó ideig párhuzamosan fog zajlani a papír- és számítógép-alapon történő adatfelvétel. Míg szükséges a korábbi eredményekkel történő összehasonlítás, addig a mérés hasonló típusú tervezése, a papír alapú teszteléstől való fokozatos eltávolodás fényében nélkülözhetetlen.

A számítógépen és papíron kitöltött teszteredmények összehasonlításával, illetve az eredményeket befolyásoló faktorok megállapításával transzformációs indexet dolgozhatunk ki. Ezek segítségével a két különböző közvetítő médián elért eredmények ekvivalenssé tehetők, ami biztosítaná a korábbi eredményekkel való összehasonlíthatóságot. A befolyásoló tényezők megállapítása után az adott részmintát ért előnyök, illetve hátrányok figyelembevételével ezek az indexek korrigálhatók.

Mindkét bemutatott vizsgálat rávilágított arra, hogy a közel azonos formátumban megjelenő tesztek jószágmutatóját, azaz az eredmények általánosíthatóságát nem befolyásolja az alkalmazott médium.

A papír- és számítógép-alapú tesztelés során adódó esetleges teljesítménykülönbségek alakulása mögött lévő okokat részletes vizsgálatokkal fel kell tárni. A tanulmányban bemutatott vizsgálatok egyöntetűen utaltak az alkalmazott feladat típusának, összetettsége befolyásoló hatásának szerepére. Minél összetettebb – extra számolást, vagy korábbi adatok megjegyzését igénylő – a feladat, annál jobban teljesítenek a diákok papír alapon, mint számítógépen, ugyanakkor a fiatalabb korosztály bátrabban válaszolt feleletválasztós kérdések esetében számítógépen, mint papír alapon. Ez a megfigyelés az idősebb minta esetében nem kimutatható.

A fiatalabbak körében tapasztalt nemek közötti esetleges teljesítménykülönbség a számítógépes tapasztalatban lévő eltérésekre vezethető vissza. Ezt igazolja, hogy mind az ötödikes fiúk, mint a főiskolás diákok feltételezhető eljutottak már arra a képességszintre ezen a területen, hogy informatikai ismereteik nem befolyásolják a jelen kutatásban alkalmazott teszt kitöltésének hatékonyságát és eredményét. Esetükben további kutatást igényel, hogy milyen mértékű eltérés, multimédiás item alkalmazás bír befolyásoló erővel.

Köszönetnyilvánítás

A tanulmány a K75274 OTKA kutatási program, az Oktatásméleti Kutatócsoport és az SZTE MTA Képességkutató Csoport keretében készült. A tanulmány írása idején Molnár Gyöngyvér Bolyai János Kutatási ösztöndíjban részesült. Köszönjük *Csizmás Edit*nek a főiskolai méréshez szükséges kérdőív adaptálásában és a mérés lebonyolításában nyújtott segítségét.

Irodalomjegyzék

Csapó Benő, Molnár Gyöngyvér és R Tóth Krisztina (2009): Comparing paper-and-pencil and online assessment of reasoning skills: A pilot study for introducing TAO in large-scale assessment in Hungary. In: Scheuermann, F. és Björnsson, J. (szerk.): *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-scale Testing*. Office for Official Publications of the European Communities, Luxemburg. 113–119.

- Csapó Benő (2001): Az induktív gondolkodás fejlődésének elemzése országos reprezentatív felmérés alapján. *Magyar Pedagógia*, 3. sz. 373–391.
- Csapó Benő, Molnár Gyöngyvér és R. Tóth Krisztina (2008): A papír alapú tesztek a számítógépes adaptív tesztekig: a pedagógiai mérés-értékelés technikájának fejlődési tendenciái. *Iskolakultúra*, 3–4. sz. 3–16.
- Farcot, M. és Latour, T. (2008): An open source and large-scale computer based assessment platform: A real winner. In: Scheuermann, F. és Pereira, A. G. (szerk.): *Towards a research agenda on computer-based assessment: Challenges and needs for European Educational Measurement*. European Commission Joint Research Centre, Ispra. 64–67.
- Józsa Krisztián (2007): *Elsajáttási motiváció*. Műszaki Könyvkiadó, Budapest.
- Józsa Krisztián és Székely Györgyi (2004): Matematikai szöveges feladatok tanítása kooperatív módszerekkel. *Magyar Pedagógia*, 104. 3. sz. 339–362.
- Jurecka, A. és Hartig, J. (2007): Computer- und netzwerkbasierter Assessment. In Hartig, J. és Klieme, E. (szerk.): *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik*. Bundesministerium für Bildung und Forschung (BMBF), Bonn, Berlin. 37–48.
- Kozma, R. (2009): Transforming education. Assessing and teaching 21st century skills. In: Scheuermann, F. és Björnsson, J. (szerk.): *The transition to computer-based assessment. New approaches to skills assessment and implications for large-scale testing*. Office for Official Publications of the European Communities, Luxemburg. 13–23.
- Martin, R. (2008): New possibilities and challenges for assessment through the use of technology. In: Scheuermann, F. és Pereira, A. G. (szerk.): *Towards a research agenda on computer-based assessment: Challenges and needs for European Educational Measurement*. European Commission Joint Research Centre, Ispra. 6–9.
- Molnár Gyöngyvér (2008): Értékelés-számítógép alapú tesztelés, online tesztkörnyezet. In: Kárpáti Andrea, Molnár Gyöngyvér, Tóth László és Föző Attila (szerk.): *A 21. század iskolája*. Nemzeti Tankönyvkiadó, Budapest. 200–211.
- OECD (2007): *PISA– The OECD Programme for International Student Assessment*. <http://www.oecd.org/dataoecd/51/27/37474503.pdf>
- Plichart P., Jadoul R., Vandenabeele L. és Latour Th. (2004): TAO, A collective distributed computer-based assessment framework built on semantic web standards. In: *Proceedings of the International Conference on Advances in Intelligent Systems – Theory and Application AISTA2004, In cooperation with IEEE Computer Society*. November 15-18, 2004. Luxembourg, Luxembourg.
- Zsolnai Anikó és Józsa Krisztián (2002): A szociális készségek kritériumorientált fejlesztésének lehetőségei. *Iskolakultúra*, 12. 4. sz. 12–20.